Generalized Difference-in-Differences

David B. Richardson, Ting Ye, and Eric J. Tchetgen Tchetgen

Presenter: Will Zhang

University of Washington

May 3, 2023

Table of contents

Motivation

Problem: to identify the causal effect of a treatment,
Difference-In-Differences (DID) Replies on "parallel trends assumption.
Confounding of the treatment effect in the pretreatment period is equivalent to confounding of the treatment effect in the post-treatment period.

Proposal:
alternative approach that can yield the identification of causal effects under different identifying conditions.

Method

Below notations are described for our interest of average effect of treatment on the treated (ATT): $\mathrm{E}\left[Y^1(i,t_1) - Y^0(i,t_1) \mid A(i)=1\right]$

- $A(i)$ denote treatment, where $A(i)=1$ if individual i is treated, $A(i)=0$ otherwise.
- $Y(i,t)$ denotes the outcome of interest for individual i at time t, where a population is observed in two periods: a pre-treatment period, $t=t_0$; and, a post-treatment period, $t=t_1$.
- $\boldsymbol{Z}(i,t)$ and $\boldsymbol{U}(i,t)$ denote measured and unmeasured variables.
- $\boldsymbol{Z}$ may denote a vector of measured variables, $Z_1,\ldots,Z_{\mathrm{p}}$, and similarly $\boldsymbol{U}$ may denote a vector of unmeasured variables $U_1,\ldots,U_Q$.
- $Y^a(i,t)$ denote individual i 's potential outcome at time t.

Method

We drop the individual argument i to focus on average causal effects. Assume:

1. Consistency for the treated, $Y(t_1) = Y^a(t_1)$, if $A = a$
2. Positivity (i.e., a small constant $\mathrm{c} > 0$, such that for any z such that $\Pr(Z = z \mid A = 1) > \mathrm{c}$ it must be that $\Pr(Z = z \mid A = 0) > \mathrm{c}$
3. No anticipation of future treatment (i.e., at t_0 individuals do not anticipate the treatment received at t_1), such that $E[Y(t_0) \mid Z] = E[Y^a(t_0) \mid Z])$, for all a.

We first describe a standard DID approach to identify the ATT. Then, we describe our proposed generalized DID

In a standard DID analysis, among all individuals the pre-treatment outcome is subtracted from the posttreatment outcome:

$$ATT_{DID} = E\left[Y\left(t_1\right) - Y\left(t_0\right) \mid A = 1\right] - \mathrm{E}\left[Y\left(t_1\right) - Y\left(t_0\right) \mid A = 0\right],$$

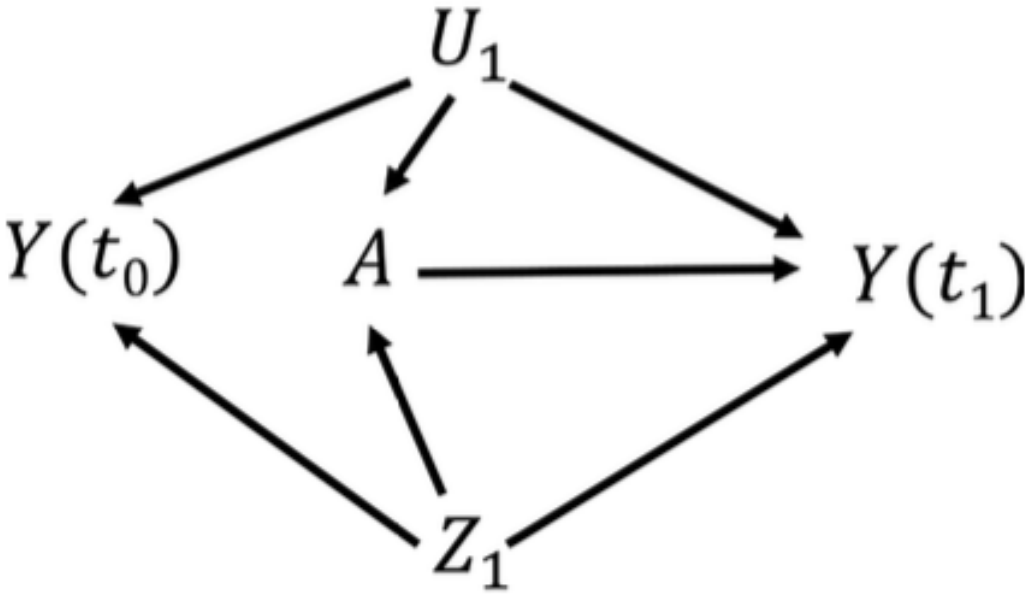
that is typically justified by conditions $1 - 3$ above as well as the parallel trends assumption,

$$\mathrm{E}\left[Y^0\left(t_1\right) - Y^0\left(t_0\right) \mid A = 1\right] = \mathrm{E}\left[Y^0\left(t_1\right) - Y^0\left(t_0\right) \mid A = 0\right],$$

such that there is no unmeasured time-varying confounders (i.e., any factor that causes a trend in the outcome over time is independent of treatment, A).

Parallel trends

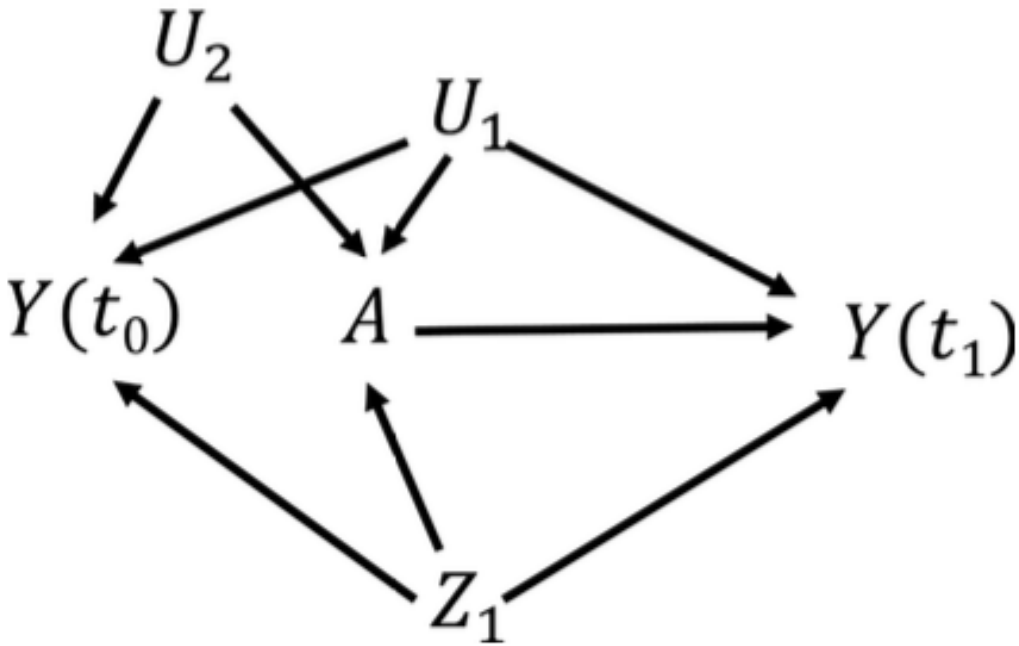
Confounding between treatment and outcome by measured or unmeasured subject-specific characteristics.



confounding bias in the posttreatment period equals to the association between the treatment and the pre-treatment outcome mean

Violation of parallel trends

Suppose that parallel trends assumption does not hold.



some unmeasured confounder of the association between treatment and pre-treatment outcomes differs in the post-treatment period.

Two stage least squares approach

Generalized DID repurposes a measured confounder, Z, of the association of interest as a bespoke instrumental variable. Define:

$$ATT_{GDID} = E\left[Y\left(t_1\right) - Y\left(t_0\right) \mid t(Z) = 1\right] - E\left[Y\left(t_1\right) - Y\left(t_0\right) \mid t(Z) = 0\right]$$

where $t(Z) = E(A \mid Z)$ and $E\left[Y\left(t_1\right) - Y\left(t_0\right) \mid t(Z)\right]$ are assumed to be linear in Z and $t(Z)$.
Stage 1: We first obtain the predicted value of A given Z,

$$\hat{A}(Z) = \hat{E}(A \mid Z)$$

by fitting a linear regression of A on Z via ordinary least squares (OLS).
Stage 2: fit a linear regression of $Y\left(t_1\right) - Y\left(t_0\right)$, on $\hat{A}(Z)$

$$\widehat{ATT}_{GDID} = \mathrm{E}\left(Y\left(t_1\right) - Y\left(t_0\right) \mid \hat{A}(Z)\right) = \hat{\beta}_0 + \hat{\beta}_1^{IV}\hat{A}(Z)$$

.

Identification

Suppose that $Z = (Z_1, Z_2)$ and instead of taking all of Z as a candidate bespoke instrumental variable:

- take Z_1 only as a bespoke instrumental variable
- Z_2 are additional measured covariates that we adjust for.

(4) Z_1 is relevant for predicting treatment: $E[A \mid Z_1, Z_2]$ depends on Z_1;
(5) No interaction between A and Z_1 in causing $Y^a(t_1)$ conditional on Z_2 and $A = 1$, such that

$$E\left[Y^{a=1}(t_1) - Y^0(t_1) \mid A = 1, Z_1, Z_2\right]$$
$$= E\left[Y^{a=1}(t_1) - Y^0(t_1) \mid A = 1, Z_1 = 0, Z_2\right]; \text{ and,}$$

Identification

(6) The additive association between Z_1 and pre-treatment outcomes is equal to the additive association between Z_1 and posttreatment outcomes (in the absence of treatment):

$$E\left[Y^0(t_0) \mid Z_1, Z_2\right] - E\left[Y^0(t_0) \mid Z_1 = 0, Z_2\right]$$
$$= E\left[Y^0(t_1) \mid Z_1, Z_2\right] - E\left[Y^0(t_1) \mid Z_1 = 0, Z_2\right]$$

Result: Under conditions 1-6 we have that for all $z_1 \neq 0$
$E\left[Y^1(t_1) - Y^0(t_1) \mid A = 1, Z_1 = z_1, Z_2\right] =$
$\frac{E[Y(t_1)-Y(t_0)|Z_1=z_1,Z_2]-E[Y(t_1)-Y(t_0)|Z_1=0,Z_2]}{E[A|Z_1=z_1,Z_2]-E[A|Z_1=0,Z_2]}$; and therefore,

$$E\left[Y^1(t_1) - Y^0(t_1) \mid A = 1\right]$$
$$= E\left[\frac{E\left[Y(t_1) - Y(t_0) \mid Z_1 = z_1, Z_2\right] - E\left[Y(t_1) - Y(t_0) \mid Z_1 = 0, Z_2\right]}{E\left[A \mid Z_1 = z_1, Z_2\right] - E\left[A \mid Z_1 = 0, Z_2\right]} \mid A =\right.$$

Under linear model specifications

$$E\left[Y^{a}\left(\mathrm{t}_1\right)-Y^{0}\left(t_1\right)\mid A=a,Z\right]=\mathrm{b}_1 a,\quad E\left[Y^{0}\left(t_1\right)\mid Z_2\right]=d_0+d_1 Z_2;$$

standard two-stage least squares approach described in the previous section, further adjusted for Z_2 in both stages, obtains a consistent estimator of $\mathrm{b}_1=E\left[Y^{1}\left(\mathrm{t}_1\right)-Y^{0}\left(t_1\right)\mid A=1\right]$.

- Condition 4 holds by definition when Z_1 is a confounder of the association of interest.
- Condition 5 is analogous to a no-interaction assumption made in the IV setting.
- Condition 6 is a Z_1 parallel trend assumption conditional on Z_2

intuition

Change in observed outcomes between t_0 and t_1 conditional on $Z_1 = z_1$ is the result of

1. change in untreated potential outcomes between t_0 and t_1
2. the causal effect of the treatment at t_1 on the treated
3. the proportion of treated units, $\mathrm{P}\left(A = 1 \mid Z_1 = z_1\right)$

Under our identifying conditions, the change in untreated potential outcomes conditional on $Z_1 = z_1$ is equal to the change in untreated potential outcomes conditional on $Z_1 = 0$. Therefore, to recover the causal effect of the treatment at t_1 for treated units at $Z_1 = z_1$ we just need to subtract off the change in observed outcomes among untreated units conditional $Z_1 = 0$ and then account for the proportion of treated units, $\mathrm{P}\left(A = 1 \mid Z_1 = z_1\right)$.

Setup

Simulated data for 1,000 studies, with 5,000 people in each study.
First scenario conformed to the "parallel trends" assumption

- Z_1 and U_1 were binary variables, taking 1 with a probability of 0.5
- A, with probability $1/\left(1+\exp\left(-\left(-0.1-0.5\times U_1+\gamma_z\times Z_1\right)\right)\right)$ of taking 1 (binary), given Z_1 is strongly ($\gamma_z = 2$), moderately ($\gamma_z = 1$), or weakly ($\gamma_z = 0.4$) associated with A.
- $Y(t=0) = (1+1\times U_1+1\times Z_1+\varepsilon)$, where $\varepsilon \sim N(0,1)$
- $Y(t=1) = (1+1\times U_1+1\times Z_1+1\times A+\varepsilon)$.

Second scenario violated the parallel trends assumption

- additional covariate U_2 from a normal $(0,1)$ distribution.
- A as a random binary variable that took a value of 1 with probability $1/\left(1+\exp\left(-\left(-0.1-0.5\times U_1-0.5\times U_2+\gamma_z\times Z_1\right)\right)\right)$.
- $Y(t=0) = (1+1\times U_1+1\times U_2+1\times Z_1+\varepsilon)$;
- $Y(t=1) = (1+1\times U_1+1\times Z_1+1\times A+\varepsilon)$.

Procedure

GDID obtain an estimate of the average change in $Y(t = 1)$ with A by a two-stage regression.

- first-stage model Z_1 was the measured variable used to predict A in linear regression
- fitted a linear regression model for $Y(t = 1) - Y(t = 0)$ as a function of the predicted value of A given Z_1.

For comparison,fit a DID with linear regression for $Y(t = 1) - Y(t = 0)$ as a function of A. We summarized results

- Monte Carlo mean and Monte Carlo standard deviation (SD)
- square root of the mean of squared difference between the estimated associations, and the specified true effect of A on Y (the root mean squared error, RMSE)
- average of standard errors (SEs)
- and coverage probability (CP) of 95% confidence intervals from normal approximation

TABLE. Monte Carlo Mean, Standard Deviation (SD), root MSE (RMSE), Average Standard Error (SE), and Coverage Probability (CP) of 95% Asymptotic Confidence Interval for 1000 Cohorts with 5,000 Observations Each. Results of Simulations of Association Between Exposure, *A*, Measured Covariate, *Z*, Unmeasured Covariate, *U*, and Outcome, *Y*

Scenario	Mean	SD	RMSE	SE	CP
Scenario 1 (Conforms to "parallel trends")					
Strong bespoke IV (mean F-statistic 1171.8)					
GDID method	1.00	0.09	0.09	0.10	95.2
Standard DID method	1.00	0.04	0.04	0.04	96.1
Moderate bespoke IV (mean F-statistic 308.5)					
GDID method	1.00	0.16	0.16	0.17	95.3
Standard DID method	1.00	0.04	0.04	0.04	96.1
Weak bespoke IV (mean F-statistic 49.7)					
GDID method	1.00	0.42	0.42	0.42	96.0
Standard DID method	1.00	0.04	0.04	0.04	94.5
Scenario 2 (Violates "parallel trends")					
Strong bespoke IV (Mean F-statistic 1067.2)					
GDID method	1.00	0.12	0.12	0.12	95.7
Standard DID method	1.39	0.05	0.39	0.05	0.0
Moderate bespoke IV (mean F-statistic 277.6)					
GDID method	1.01	0.21	0.21	0.22	95.7
Standard DID method	1.44	0.05	0.44	0.05	0.0
Weak bespoke IV (Mean F-statistic 44.0)					
GDID method	1.01	0.55	0.55	0.55	96.5
Standard DID method	1.46	0.05	0.47	0.05	0.0

Empirical Results Omitted

Thank you